PRINCIPLES OF MEDICAL STATISTICS

BRADFORD HILL

SEVENTH EDITION

PRINCIPLES OF MEDICAL STATISTICS

BY

AUSTIN BRADFORD HILL

C.B.E., D.Sc., PH.D., F.R.S.

Professor Emeritus of Medical Statistics in the University of London; Past Honorary Director of the Statistical Research Unit of the Medical Research Council; Past President of the Royal Statistical Society; Civil Consultant in Medical Statistics to the Royal Navy and the Royal Air Force; Member of the International Statistical Institute

SEVENTH EDITION

REVISED AND ENLARGED

PUBLISHED BY

THE LANCET LIMITED

7 ADAM STREET ADELPHI LONDON

1961

First Edition, 1937
Second Edition, 1939
Third Edition, 1942
Third Edition Reprinted, 1945
Third Edition Reprinted, 1946
Fourth Edition, 1948
Fourth Edition Reprinted, 1949
Fifth Edition, 1950
Fifth Edition Reprinted, 1952
Sixth Edition, 1955
Sixth Edition Reprinted, 1959
Seventh Edition, 1961

Translated into
Spanish, *First Edition*, 1950
Second Edition, 1958
Russian, 1958
Korean, 1960
Polish, 1962

To
F. M. H.

PRINTED IN GREAT BRITAIN
BY R. & R. CLARK, LTD., EDINBURGH

FOREWORD TO FIRST EDITION

In clinical medicine to-day there is a growing demand for adequate proof of the efficacy of this or that form of treatment. Often proof can come only by means of a collection of records of clinical trials devised on such a scale and in such a form that statistically reliable conclusions can be drawn from them. However great may be our aversion to figures, we cannot escape the conclusion that the solution of most of the problems of clinical or preventive medicine must ultimately depend on them. Even those who pretend to despise this method of approach find that any assessment of success or failure which is based on fact rather than on opinion must nearly always be expressed in some numerical form—*e.g.* when the medical observer reports that he has treated so many cases with a favourable result in such and such a proportion, or the public health worker assesses the attack-rate on a population inoculated against some infection. But often, unfortunately, the figures used are either insufficient in number or documentation or too limited in their scope to bear the weight of the interpretation that is placed upon them. An additional difficulty is that few medical men have been trained to interpret figures or to analyse and test their meaning by even an elementary statistical technique. We have reason to believe that there is now a steadily increasing demand among both clinical and public health workers for some knowledge of that technique and a realisation that it is not much good collecting figures more or less haphazardly and then to expect a professional statistician to draw conclusions from them.

Acting upon this belief, last year we invited Dr. A. Bradford Hill to prepare for our columns a series of short simple articles on such methods as his experience of medical statistics had shown him would be most useful in that

field. He has succeeded beyond our hopes in demonstrating some of the ways in which investigations can be planned and figures derived from them can be analysed in order to yield fruitful results. He has chosen examples from medical publications to illustrate both the types of problem with which the medical worker is faced and the kinds of error he is most liable to make; and wherever possible he has either avoided mathematical presentation of his material or has led up to it so skilfully as to rob equations and even square roots of their traditional horrors. This consideration for those who are not mathematically minded has resulted in an exceptionally clear exposition of a difficult subject. We are happy to accede to the many requests we have received for the reissue of these articles in book form.

EDITOR OF "THE LANCET"

June 1937

PREFACE TO SEVENTH EDITION

Two long sea voyages provided me with an opportunity to give a great deal of uninterrupted reflection to this book and, as a result, to make considerable changes in it. I have added (I fear, the consequence of *too* much leisure) three new chapters, namely Chapters II and IV concerned with the elements of sampling and the carrying out of scientific investigations, and Chapter XIX dealing with the problems of defining and measuring sickness. In some compensation for this increase I have reduced the space previously given to the standardisation of death-rates and have incorporated its customary rates and indices in one chapter (XVII) instead of the previous two.

Other major revisions that I have thought it wise to make include an extension of the section on graphs and diagrams (Chapter V), the introduction of the "t" test (Chapter XII) and of a χ^2 test of the trend in a series of percentages (Chapter XIV), and a rather more detailed account of the construction of a life table as a means of observing and analysing the follow-up of a group (Chapter XVIII). Additions to Chapter XX on Clinical Trials comprise a discussion of trials made "within," rather than "between," patients, and an emphasis on the importance of being wary of subsequent exclusions of patients, for any reason, from originally randomly constructed groups. I have also taken pleasure in adding a few more "fallacies and difficulties" (Chapters XXI to XXIII) and in rearranging the examples in those three chapters in a more logical order.

While making revisions on this scale I have also taken the opportunity to alter the order of some of the chapters, which, with succeeding editions, had become awry; and throughout the book I have rewritten many phrases and sentences. With the increasing use of statistics in medicine

it is my hope that these many changes will not offend old friends and will enable the book to continue to aid newcomers to the field—and to continue to aid them as much towards clear ways of thinking as in the use of the simple methodology that it expounds.

I am greatly indebted to Dr. P. Armitage and Mr. I. D. Hill for reading my expanded text and for their advice and criticisms. Contrary to custom, for the faults that remain I trust sincerely that the reader may hold them largely responsible. The values in the table of "t" I extracted from the Biometrika Tables for Statisticians and I am grateful to the Biometrika Trustees for permission to do so.

A. BRADFORD HILL

January 1961

PREFACE TO FIRST EDITION

"**Statistics are curious things.** They afford one of the few examples in which the use, or abuse, of mathematical methods tends to induce a strong emotional reaction in non-mathematical minds. This is because statisticians apply, to problems in which we are interested, a technique which we do not understand. It is exasperating, when we have studied a problem by methods that we have spent laborious years in mastering, to find our conclusions questioned, and perhaps refuted, by someone who could not have made the observations himself. It requires more equanimity than most of us possess to acknowledge that the fault is in ourselves."

THESE are the opening words of a leading article which introduced the chapters of this book to the readers of *The Lancet*, where they appeared weekly during the first four months of 1937. As a statistician I may be permitted to view the problem the other way round. No statistician (in spite of views to the contrary) particularly enjoys refuting other workers' conclusions, especially when he knows that much patience and much hard work have been expended in the collection of data. He has no burning desire to be called in as an expert witness at the conclusion of the case for the plaintiff (on a subject-matter he may not fully understand) to say that this is "significant" and that is, statistically speaking, nonsense. It even requires some equanimity to be unmoved when he is transferred to the dock on a charge of being merely an armchair critic or a confirmed sceptic. In other words, neither party to the transaction is particularly happy.

There seems to me to be only one way of escape. The worker in medical problems, in the field of clinical as well as preventive medicine, must *himself* know something of statistical technique, both in experimental arrangements and in the interpretation of figures. To enable him to acquire some knowledge of this technique I have tried to set down as simply as possible the statistical methods that experience has shown me to be most helpful in the problems with which medical workers are concerned. I have used examples taken from medical inquiries

in the attempt to make clear these methods of analysis, and have sought to show by illustration where and why workers make mistakes in their interpretation of figures. I know that I have been guilty of some repetition; my excuse is the repetition in published papers of those elementary statistical errors which a very little knowledge of statistics would be sufficient to prevent.

I am much indebted to Professor Major Greenwood, F.R.S., Professor W. W. C. Topley, F.R.S., Dr. J. O. Irwin, and Mr. W. T. Russell, for reading the whole or part of the manuscript or proofs. I have gained much from their criticisms. The original articles were written for *The Lancet* at the suggestion of Dr. M. H. Kettle, and I owe a very great deal to her encouragement and interest.

The table of χ^2 is reproduced by kind permission of the author and publishers from *Statistical Methods for Research Workers*, by Professor R. A. Fisher, Sc.D., F.R.S. (6th edition, 1936, Oliver & Boyd, Edinburgh and London).

A. BRADFORD HILL

June 1937

CONTENTS

I

THE AIM OF THE STATISTICAL METHOD

"Is the application of the numerical method to the subject-matter of medicine a trivial and time-wasting ingenuity as some hold, or is it an important stage in the development of our art, as others proclaim?" Whatever may have been the reactions of medically qualified readers to that question, propounded by the writer of an article on medical statistics in 1921 (*Lancet*, 1, 985), it must be admitted that in subsequent years there has been a continual and substantial increase in the number of papers contributed to medical journals of which the essence is largely statistical. Not only has there developed an enhanced knowledge of, and respect for, the national registers of life and death which the Registrars-General of the United Kingdom (and similar authorities in other countries) annually publish and analyse, but there is an increasing number of workers who endeavour to apply numerical methods of analysis to their records obtained, by observation or experiment, in clinical and other branches of medicine. Many such workers, however, have had little or no training in statistical method, and many of them find the more mathematical methods of the professional statistician, as has been said, "obscure and even repellent." Often enough, indeed, the argument is put forward that the use of such mathematical methods is quite unjustifiable, that the accuracy of the original material is not sufficient to bear the weight of the treatment meted out to it. This assertion is not strictly logical. If a collection of figures is worth a statistical analysis at all, it is, obviously, worth the best form of statistical analysis—*i.e.* the form which allows the maximum amount of information to be derived from the data.

Whether mathematical statistical methods *are* the best form in particular cases, whether they are essential or may be regarded as an unnecessary elaboration must turn rather upon this question : Can we in any of the problems of medical statistics reach satisfactory results by means of relatively simple numerical methods only ? In other words : Can we satisfactorily test hypotheses and draw deductions from data that have been analysed by means of such simple methods ? The answer is undoubtedly yes, that many of the figures included today in medical papers can by relatively simple statistical methods be made to yield information of value, that where the yield is rather less than that which might be obtained by more erudite methods which are not at the worker's command the best should not be made the enemy of the good, and that even the simplest statistical analysis carried out logically and carefully is an aid to clear thinking with regard to the meaning and limitations of the original records. If these conclusions are accepted, the question immediately at issue becomes this : Are simple methods of the interpretation of figures only a synonym for common sense or do they involve an art or knowledge which can be imparted ? Familiarity with medical statistics leads inevitably to the conclusion that common sense is *not* enough. Mistakes which when pointed out look extremely foolish are quite frequently made by intelligent persons, and the same mistakes, or types of mistakes, continue to crop up again and again. There is often lacking what has been called a "statistical tact, which is rather more than simple good sense." That tact the majority of persons must acquire (with a minority it is undoubtedly innate) by a study of the basic principles of statistical thought and method.

The object of this book is to discuss these basic principles in an elementary way and to show, by representative examples taken from medical literature, how these principles are frequently forgotten or ignored. There is no doubt that the discussion will often appear too simple and that some of the mistakes to which space is given will be thought too futile to need attention. That such is not the case is revealed by the recurrence of these mistakes and the neglect of these

elementary principles, a feature with which every professional statistician is familiar in the papers submitted to him by their authors for " counsel's opinion."

Definition of Statistics

Whereas the laboratory worker can frequently exclude variables in which he is not interested and confine his attention to one or more controlled factors at a time, the worker in clinical or preventive medicine is often unable to experiment and must inevitably use records which may be influenced by factors which he cannot control but have essentially to be taken into account. The essence of the statistical method lies in the elucidation of the effects of these multiple causes. By statistics, therefore, we mean " quantitative data affected to a marked extent by a multiplicity of causes," and by statistical method " methods specially adapted to the elucidation of quantitative data affected by a multiplicity of causes " (Yule and Kendall, *An Introduction to the Theory of Statistics*, Charles Griffin & Co., Ltd., London).

For example, suppose we have a number of children all of whom have been in contact with measles and to a proportion of them is given an injection of gamma globulin. We wish to know whether the treatment prevents the development of a clinical attack. It is possible that the risk of developing an attack is influenced by age, by sex, by social class and all that that denotes, by duration and intimacy of contact, by general state of health. A statistical analysis necessitates attention to *all* these possible influences. We must endeavour to equalise the groups we compare in every possibly influential respect except in the one factor at issue —namely, the treatment. If we have been unable to equalise the groups *ab initio* we must equalise them to the utmost extent by the mode of analysis. As far as possible it is clear, however, that we should endeavour to eliminate, or allow for, these extraneous or disturbing causes when the observations are planned ; with such planning maybe we can determine not only whether the treatment is of value but whether it is more efficacious at one age than another, etc. It is a

serious mistake to rely upon the statistical method to eliminate disturbing factors at the completion of the work. *No* statistical method can compensate for badly planned observations or for a badly planned experiment.

Planning and Interpretation of Experiments

It follows that the statistician may be able to advise upon the statistical lines an experiment such as that referred to above should follow. Elaborate experiments can be planned in which quite a number of factors can be taken into account statistically at the same time (see, for example, *Experimental Designs* by W. G. Cochran and G. M. Cox and *Planning of Experiments* by D. R. Cox). It is not the intention to discuss these more complex methods of planning and analysis; attention is mainly confined to the simpler types of experimental arrangement with which medical workers are familiar. Limitation of the discussion to that type must not be taken to mean that it is the best form of experiment in a particular case.

The essence of the problem in a simple experiment is, as emphasised above, to ensure beforehand that, as far as is possible, the control and treated groups are the same in all *relevant* respects. The word "relevant" needs emphasis for two reasons. First, it is obvious that no statistician, when appealed to for help, can be aware of all the factors that are, or may be, relevant in particular medical problems. From general experience he may well be able to suggest certain broad disturbing causes which should be considered in planning the experiment (such as age and sex in the example above), but with factors which are narrowly specific to a particular problem he cannot be expected to be familiar. The onus of knowing what is likely to be relevant in a specific problem must rest upon the experimenter, who is, presumably, familiar with that narrow field. Thus, when the statistician's help is required it is his task to suggest means of allowing for the disturbing causes, either in planning the experiment or in analysing the results, and not, as a rule, to determine what *are* the relevant disturbing causes. At

the same time successful collaboration demands that the statistician learn all he can of the problem at issue and the experimenter (clinician, medical officer of health, etc.) all he can of the statistical approach. Without substantial knowledge on both sides the blind may well lead the blind.

The second point that must be observed as regards the equality of groups in all relevant respects is the caution that must attend the interpretation of statistical results. If we find that Group A differs from Group B in some characteristic, say, its mortality-rate, can we be certain that that difference is due to the fact that Group A was inoculated (for example) and Group B was uninoculated? Are we certain that Group A does not differ from Group B in some other character relevant to the issues as well as in the presence or absence of inoculation? For instance, in a particular case, inoculated persons might, on the average, belong to a higher social class than the uninoculated and therefore live in surroundings in which the risk of infection was less. We can never be *certain* that we have not overlooked some relevant factor or that some factor is not present which could not be foreseen or identified. It is because he knows a complex chain of causation is so often involved that the statistician is, as it appears to many persons, an unduly cautious and sceptical individual.

Statistics in Clinical Medicine

The essence of an experiment in the treatment of a disease lies in comparison. To the dictum of Helmholtz that "all science is measurement," we should add, Sir Henry Dale has pointed out, a further clause, that "all true measurement is essentially comparative." On the other hand there is a common catch-phrase that human beings are too variable to allow of the contrasts inherent in a controlled trial of a remedy. Yet if each patient is "unique" it is difficult to see how any basis for treatment can be sought in the past observations of other patients—upon which clinical medicine is founded. In fact, of course, physicians must, and do, base their "treatment of choice" upon what they have seen

happen before—whether it be in only two or three cases or in a hundred.

However, though, broadly speaking, human beings are not unique in their responses to some given treatment, there is no doubt that they are likely to be variable, and sometimes extremely variable. Two or three observations may, therefore, give, merely through the customary play of chance, a favourable picture in the hands of one doctor, an unfavourable picture in the hands of another. As a result, the medical journals become an arena for conflicting claims—each in itself, maybe, perfectly true of what the doctor saw but insufficient to bear the weight of the generalisation placed upon it.

Far, therefore, from arguing that the statistical approach is impossible in the face of human variability, we must realise that it is *because* of variability that it is often essential. It does not follow, to meet another common criticism, that it invariably demands large numbers. It may do so; it depends upon the problem. But the responses to treatment of a single patient are clearly a statement of fact—so far as the observations were truly made and accurately recorded. Indeed that single case may give, in certain circumstances, evidence of vital importance.

If, for example, we were to use a new drug in a proved case of acute leukaemia and the patient made an immediate and indisputable recovery, we should have a result of the most profound importance. The reason underlying our acceptance of merely one patient as illustrating a remarkable event—not necessarily of cause and effect—is that long and wide experience has shown that in their response to acute leukaemia human beings are *not* variable. They one and all fail to make immediate and indisputable recoveries. They one and all die. Therefore, although it would clearly be most unwise upon one case to pass from the particular to the general, it would be sheer madness not to accept the evidence presented by it.

If, on the other hand, the drug were given to a patient suffering from acute rheumatic fever and the patient made an immediate and indisputable recovery, we have little basis

for remark. That recovery may clearly have followed the administration of the drug without the slightest probability of a related cause and effect. With this disease human beings *are* variable in their reactions—some may die, some may have prolonged illnesses but recover eventually with or without permanent damage, some may make immediate and indisputable recoveries—whatever treatment we give them. We must, therefore, have more cases before we can reasonably draw inferences about cause and effect. We need a statistical approach and a designed experiment (the details are discussed in Chapter XX).

While, therefore, in many instances we do need larger numbers for a sound assessment of a situation, it certainly does not follow — as is sometimes asserted — that the statistician would have rejected some of the original and fundamental observations in medicine on the grounds of their small number. To take a specific example, *fragilitas ossium* was originally described on two cases and this, a later writer said, statisticians would regard as useless evidence. But why should they ? If exact descriptions and illustrations were given of these two cases, then, of course, they form part of the body of scientific knowledge. They are undeniable evidence of an occurrence. What *can* happen, what *does* exist, quite regardless of the *frequency* of occurrence and irrespective of causation or association, may be observed, as already stated, even on a sample of one. It can only be in relation to an appeal from the particular to the general that a statistician—and, equally, any trained scientific worker—could object. If on the basis of the two cases the clinician, in practice, let us say, near the London meat market, should argue that the condition was specific to butchers, then one might suggest that the experience was too limited in size and area to justify any such generalisation.

In short, there is, and can be, no magic number for either clinician or statistician. Whether we need one, a hundred, or a thousand observations turns upon the setting of our problem and the inferences that we wish to draw.

It must be clear, too, that almost without statistics, and certainly without accurate measurement, the mental, or